Lecture 10.

Syndrome decoding algorithm.

How do we find a closest element in C? Go over all codewords? There is a more efficient algorithm.

Let C be an $[n,k]$-code in F^n. In particular, C is a subgroup in the abelian group F^n.

Cosets : $C + a = \{v + a \mid v \in C\}$.

Recall the main properties of cosets:

(1) Two cosets are either disjoint or identical,

(2) $C + a = C + b \iff a - b \in C$

(3) $|C + a| = |C|$

(4) F^n = disjoint union of cosets,

$F^n = (C + a_1) \dot\cup (C + a_2) \dot\cup \cdots \dot\cup (C + a_2)$.

Here $r = |F^n : C|$ is the index of C, $r =$

$$\frac{|F^n|}{|C|} = \frac{q^n}{q^k} = q^{n-k}.$$

<u>Definition</u>. A vector of minimal weight in a coset $C+a$ is called a <u>coset leader</u>. There may be several leaders in the same coset.

<u>Example</u>. C is a $[4,2]$-binary code with generator matrix

$$G = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \end{pmatrix},$$

$|\mathbb{Z}(2)^4| = 16$, $|C| = 2^2 = 4$, we have 4 cosets.

$C = \{\underline{0}, (1011), (0101), (1110)\}$

$C + (1000) = \{\underline{(1000)}, (0011), (1101), (0110)\}$

$C + (0100) = \{\underline{(0100)}, (1111), \underline{(0001)}, (1010)\}$

$C + (0010) = \{\underline{(0010)}, (1001), (0111), (1100)\}$

The leaders in the cosets above are underlined. Note that the 3d coset has two leaders.

Let C be an $[n,k]$-code, let H be the parity check matrix of C, $H = [\underbrace{\quad\quad}_{n}]\, n-k$.

<u>Definition</u>. For a vector $a \in F^n$ the <u>syndrome</u> $s(a)$ of the vector a is defined as

$$s(a) = H a^T$$

It is a vector of length $n-k$.

<u>Lemma</u>. All vectors in a given coset have the same syndrome.

<u>Proof</u>. Let v, w lie in the same coset. Then $v - w \in C$. Then $H(v-w)^T = 0$, which implies

$$H v^T = H w^T$$

↓

Let's find the parity check matrix for the [4,2]-code above

$$\begin{cases} x_1 + x_3 + x_4 = 0 \\ x_2 + x_4 = 0 \end{cases} \qquad \begin{aligned} x_1 &= -x_3 - x_4 \\ x_2 &= -x_4 \end{aligned}$$

$x_3 = 1,\ x_4 = 0,\ x_1 = 1,\ x_2 = 0:\ (1\,0\,1\,0)$

$x_3 = 0,\ x_4 = 1,\ x_1 = 1,\ x_2 = 1:\ (1\,1\,0\,1)$

$$H = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \end{pmatrix}$$

Syndromes of the coset leaders:

Leader	Syndrome
~~(1000)~~ $\mathbf{0}$	$\begin{pmatrix} 0 \\ 0 \end{pmatrix}$
(1000)	$\begin{pmatrix} 1 \\ 1 \end{pmatrix}$
(0001)	$\begin{pmatrix} 0 \\ 1 \end{pmatrix}$
(0010)	$\begin{pmatrix} 1 \\ 0 \end{pmatrix}$

Suppose that a vector $v \in C$ has been sent and a vector w has been received. The vector $e = w - v$ is called the <u>error vector</u>

$wt(e) = \#$ of errors $= \#$ of coordinates that have been altered.

Clearly, $Hw^T = He^T$ since $v \in C$ and $Hv^T = 0$.

<u>Syndrome Decoding Algorithm.</u>

Suppose that we have received a vector $w \in F^n$.

1) Compute the syndrome
$$S(w) = Hw^T \; ;$$
2) find a coset leader e having the same syndrome as w;
3) if e is the only leader in the coset with syndrome $S(w)$ then

decode $v = w - e$. If there are ≥ 2 leaders then conclude that w can not be corrected with the code C.

This is a reformulation of the closest vector method.

<u>Advantage</u>: we can compute leaders and their syndromes in advance.

If the channel makes $\leq [\frac{d-1}{2}]$ errors then this method always works.

For the $[4,2]$ example above $d = 2$,

$[\frac{d-1}{2}] = 0$. Zero strength.

But $(1\,0\,0\,0)$ and $(0\,0\,1\,0)$ are single leaders in their cosets.

Therefore the code can correct single errors in the 1st and the 3d coordinates.

Example: Hamming Code.

Lexicographical Comparison. Suppose that we have two 0,1-vectors of the same length n: $v = \begin{pmatrix} i_1 \\ \vdots \\ i_n \end{pmatrix}$ and $w = \begin{pmatrix} j_1 \\ \vdots \\ j_n \end{pmatrix}$. Start with i_1 and j_1. If $i_1 = 1, j_1 = 0$ then $v > w$. If $i_1 = 0$, $j_1 = 1$ then $w > v$. If $i_1 = j_1$ then forget about the first components and move on to i_2, j_2. If $i_2 = 1, j_2 = 0$ then $v > w$. If $i_2 = 0, j_2 = 1$ then $v < w$. If $i_2 = j_2$ then move on to the 3d components, and so on.

Let C be a $[7,4]$-code with the parity matrix $H = \underbrace{[\qquad]}_{7}\}3$ having as columns all nonzero 0,1-vectors of length 3 (there are 7 such vectors, $7 = 2^3 - 1$), ordere lexicographically,

$$H = \begin{pmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{pmatrix}.$$

Any two columns are linearly independent, but the first 3 columns are linearly dependent. Hence $d = 3$. The code can correct single errors as $\left[\frac{3-1}{2}\right] = 1$. There are $2^{7-4} = 8$ cosets.

Let $e_i = (0 \cdots \underset{i}{1} \, 0 \cdots 0)$. The vectors $0, e_1, e_2, \dots, e_7$ are representatives of different cosets. Indeed, $e_i - e_j \notin C$ as $d(C) = 3$. The zero vector and these vectors of weight 1 are the leaders in their cosets, unique leaders. The syndrome of e_i is $S(e_i) = He_i^T =$ the i-th column of H.

Now the algorithm is:

having received a vector w compute $S(w) = Hw^T$. If $S(w) = 0$ then no error has occured. Otherwise $S(w)$ = the i-th column of H. To get the vector v replace the i-th coordinate in w.

Sphere-packing.

Let C be an $[n,k,d]$-code, $|C| = q^k$.

For fixed $|C|$ and d we want to minimize n.

For fixed n and d we want to maximize $|C|$.

Theorem (bounding $|C|$ in terms of n and d):

$$|C| \leq \frac{q^n}{\sum_{i=0}^{[d-1/2]} \binom{n}{i} (q-1)^i}$$

Proof. The volume (number of elements) of a ball of radius $r = [\frac{d-1}{2}]$ is

$$\sum_{i=0}^{r} \binom{n}{i}(q-1)^i$$

Balls of radius r with centers at different points of C do not intersect. At total we have q^n points. This implies the inequality. □

Def. A code C such that $|C| = \frac{q^n}{|B(0,r)|}$,

or, in other words, F^n is the disjoint union of balls $F = \dot{\bigcup}_{v \in C} B(v, r)$ is called perfect.

Example (trivial). $C = F^n$. Hence $d = 1$,

$[\frac{d-1}{2}] = 0$.

<u>Example</u> (also trivial). $n = 2m+1$, C consists of two vectors: 0 and $(\underbrace{11\cdots1}_{n})$. Here $d = n = 2m+1$, $\frac{d-1}{2} = m$. The volume of the ball is $\sum_{i=0}^{m}\binom{n}{i} = \frac{1}{2}\sum_{i=0}^{n}\binom{n}{i} = \frac{1}{2}2^n = 2^{n-1}$. We have $|C| = 2 = \frac{2^n}{2^{n-1}}$.

<u>Hamming Codes</u>. Let $n = 2^{m-1}$ $(m \geq 2)$. We will start with a parity check matrix H, $H = \underbrace{\left[\quad\quad\right]}_{n}\Big\}m$, columns are all the distinct nonzero $0,1$-vectors in the lexicographical order. The order is not so important, they are all equivalent up to permutation equivalence.

the code C has dimension $k = n - m = 2^m - m - 1$. Any two different columns of H are linearly independent. There are 3 columns that are dependent. Hence $d = 3$, $\left[\frac{d-1}{2}\right] = 1$. The ball of radius 1 has volume $1 + n = 2^m$. Now,

$$\frac{2^n}{2^m} = 2^{n-m}.$$

The code C has $2^k = 2^{n-m}$ elements. Hence the code is perfect.

<u>Hamming codes over arbitrary finite fields.</u>

Let $|F| = q$. Let $n = \frac{q^m - 1}{q - 1}$, $m \geq 2$. There are exactly n distinct 1-dimensional subspaces of F^m (Exercise: prove it). Let

$$H = \underbrace{\left[\quad \quad \quad \right]}_{n} \Big\} m,$$

the columns are representatives of all distinct 1-dimensional subspaces of F^m.

Again $d=3$.

<u>Exercise</u>: prove that it is a perfect code.

<u>Examples of Nonlinear Codes</u> (<u>Vasil'ev Codes</u>)

<u>Exercise</u>. If C is a perfect (not necessarily linear) code then $\forall a \in F^n$ $C+a$ is again a perfect code.

We will introduce a construction

a binary perfect code with $d=3$ of length n

$\Rightarrow$ a binary perfect code of length $2n+1$,

that is not linear and not a coset of a linear code, $d=3$.

Let E be a perfect code with $d=3$,

$n=2^m-1$, containing 0 (for example the Hamming code $H_m(2)$).

Let $f : E \to \mathbb{Z}(2)$ be a function,

$f(0)=0$.

Let $\pi : \mathbb{Z}(2)^n \to \mathbb{Z}(2)$, $\pi((\alpha_1, \dots, \alpha_n)) = \alpha_1 + \dots + \alpha_n$.

Let $C \subset \mathbb{Z}(2)^{2n+1}$ be defined as follows:

$$C = \{ (v, v+a, \pi(v) + f(a) \mid v \in \mathbb{Z}(2)^n, a \in E \},$$

$$C \subseteq \mathbb{Z}(2)^{2n+1}.$$

<u>Theorem</u>: C is a perfect code with $d=3$. If f is not linear then C is not linear and not a coset of a linear code.

<u>Proof</u>. Let us show that the minimal distance of the code C is 3. Let $x = (v, v+a, \pi(v) + f(a))$, $y = (w, w+b, \pi(w) + f(b))$.

Then $d(x,y) = d(v,w) + d(v+a, w+b) + d(\pi(v)+f(a), \pi(w)+f(b))$

We will consider all possible cases:

1) $u = w$. Then $x \neq y \Leftrightarrow a \neq b$. Then $d(v+a, v+b) = d(a,b) \geq 3$.

~~2) 2.1) $u \neq w$, $d(u,w) = 1$; $a = b$. Then~~

2.1) $u \neq w$ and, moreover, $d(u,w) \geq 2$; $a=b$

Then $d(v,w) \geq 2$, $d(v+a, w+a) = d(v,w) \geq 2$, hence

$d(x,y) \geq 4$;

2.2) $u \neq w$, $d(u,w) = 1$, $a=b$. Then $d(v,w)$

$= d(v+a, w+a) = 1$. Since $d(u,w)=1$ it

follows that $\pi(v) \neq \pi(w)$, hence

$d(\pi(v) + f(a), \pi(w) + f(a)) \geq 1$ and again

$d(x,y) \geq 3$.

3) $u \neq w$, $a \neq b$. We have

$$d(v,w) + d(v+a, w+b) = d(v,w) +$$

$$d(v+a-b, w) \underset{\substack{\uparrow \\ \text{triangle inequality}}}{\geq} d(v, v+a-b) =$$

$$= d(v+b, v+a) = d(b,a) \geq 3.$$

We proved that $d \geq 3$.

Now as in 2.2 let $a=b$, $u=e_1=(1,\underbrace{0\cdots0}_{n-1})$, $v=0$.

Then $d(x,y)=3$.

Let us show that the code C is perfect.

$|E| = \dfrac{2^n}{2^m} = 2^{n-m}$ since the code E is perfect

Now, $|C|=|\mathbb{Z}(2)^n|\cdot|E| = 2^n\cdot 2^{n-m} = 2^{2n-m}$.

The volume of the ball of radius 1 in $\mathbb{Z}(2)^{2n+1}$ is $2n+2 = 2(n+1) = 2\cdot 2^m = 2^{m+1}$.

Finally,
$$2^{2n-m} = \frac{2^{2n+1}}{2^{m+1}},$$

the code C is perfect.

The code C contains the zero vector 0. Hence it is either linear or not a coset of a linear code. Let us show that C is not linear ~~unle~~ if f is not linear.

We have

$x+y=(u+v,\ u+v+a+b,\ \pi(u)+\pi(v)+f(a)+f(b))$.

If E is not linear then we can choose $a,b\in E$ such that $a+b\notin E$. Then $x+y\notin C$.

Let E be linear. Since $\pi(u)+\pi(v)=\pi(u+v)$ the vector $x+y$ lies in C if and only if $f(a)+f(b)=f(a+b)$. $\downarrow$